

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- 5 (A) NAME: SOLVAY (Société Anonyme)
(B) STREET: rue du Prince Albert, 33
(C) CITY: Brussels
(E) COUNTRY: Belgium
(F) POSTAL CODE: 1050
(G) TELEPHONE: (02) 509.61.11

10 (ii) TITLE OF INVENTION: Xylanase, microorganisms producing it, DNA molecule [sic], methods for preparing this xylanase and uses of the latter

(iii) NUMBER OF SEQUENCES: 29

(iv) COMPUTER READABLE FORM:

- 15 (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS: MS-DOS

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 663 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAAATCGTCA CCGACAATTC CATTGGCAAC CACGATGGCT ATGATTATGA ATTTTGAA

GATAGCGGTG GCTCTGGGAC AATGATTCTC AATCATGGCG GTACGTTCAAG TGCCCAATGG 120
ACAATGTTA ACAACATATT ATTCCGTAAA GGTAaaaaAT TCAATGAAAC ACAAACACAC 180
CAACAAGTTG GTAACATGTC CATAAACTAC GGAGCCAAct TCCAACCAAA TGGTAATGCG 240
TATTTATGCG TCTATGGTTG GACTGTTGAC CCTCTTGTGCG AATATTATAT TGTCGACAGT 300
TGGGGCAACT GGCGTCCACC AGGAGCAACG CCTAAGGGGA CCATCACTGT TGATGGAGGA 360
ACATATGATA TCTACGAGAC TCTTAGAGTC AATCAACCCT CCATTAAGGG GATTGCCACA 420
TTTAAACAAT ATTGGAGTGT TCGAAGATCG AAACGCACGA GTGGCACGAT TTCTGTCAGC 480
AACCACTTA GAGCGTGGGA AAACCTAGGG ATGAATATGG GGAAAATGTA TGAAGTCGCG 540
CTTACTGTAG AAGGCTATCA AAGTAGCGGA AGTGCTAATG TATATAGCAA TACACTAAGA 600
ATTAACGGTA ACCCTCTCTC AACTATTAGT AATGACGAGA GCATAACTTT GGATAAAAAC 660
AAT 663

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 663 base pairs**
- (B) TYPE: nucleic acid**
- 5 (C) STRANDEDNESS: single**
- (D) TOPOLOGY: linear**

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CAA ATC GTC ACC GAC AAT TCC ATT GGC AAC CAC GAT GGC TAT GAT TAT			48
Gln Ile Val Thr Asp Asn Ser Ile Gly Asn His Asp Gly Tyr Asp Tyr			
1	5	10	15
GAA TTT TGG AAA GAT AGC GGT GGC TCT GGG ACA ATG ATT CTC AAT CAT			96
Glu Phe Trp Lys Asp Ser Gly Gly Ser Gly Thr Met Ile Leu Asn His			
20	25	30	
GGC GGT ACG TTC AGT GCC CAA TGG AAC AAT GTT AAC AAC ATA TTA TTC			144
Gly Gly Thr Phe Ser Ala Gln Trp Asn Asn Val Asn Asn Ile Leu Phe			
35	40	45	
CGT AAA GGT AAA AAA TTC AAT GAA ACA CAA ACA CAC CAA CAA GTT GGT			192
Arg Lys Gly Lys Lys Phe Asn Glu Thr Gln Thr His Gln Gln Val Gly			
50	55	60	
AAC ATG TCC ATA AAC TAC GGA GCC AAC TTC CAA CCA AAT GGT AAT GCG			240
Asn Met Ser Ile Asn Tyr Gly Ala Asn Phe Gln Pro Asn Gly Asn Ala			
65	70	75	80
TAT TTA TGC GTC TAT GGT TGG ACT GTT GAC CCT CTT GTC GAA TAT TAT			288
Tyr Leu Cys Val Tyr Gly Trp Thr Val Asp Pro Leu Val Glu Tyr Tyr			
85	90	95	
ATT GTC GAC AGT TGG GGC AAC TGG CGT CCA CCA GGA GCA ACG CCT AAG			336
Ile Val Asp Ser Trp Gly Asn Trp Arg Pro Pro Gly Ala Thr Pro Lys			
100	105	110	
GGG ACC ATC ACT GTT GAT GGA GGA ACA TAT GAT ATC TAC GAG ACT CTT			384
Gly Thr Ile Thr Val Asp Gly Gly Thr Tyr Asp Ile Tyr Glu Thr Leu			
115	120	125	
AGA GTC AAT CAA CCC TCC ATT AAG GGG ATT GCC ACA TTT AAA CAA TAT			432
Arg Val Asn Gln Pro Ser Ile Lys Gly Ile Ala Thr Phe Lys Gln Tyr			
130	135	140	

TGG AGT GTT CGA AGA TCG AAA CGC ACG AGT GGC ACG ATT TCT GTC AGC	480		
Trp Ser Val Arg Arg Ser Lys Arg Thr Ser Gly Thr Ile Ser Val Ser			
145	150	155	160
AAC CAC TTT AGA GCG TGG GAA AAC TTA GGG ATG AAT ATG GGG AAA ATG	528		
Asn His Phe Arg Ala Trp Glu Asn Leu Gly Met Asn Met Gly Lys Met			
165	170	175	
TAT GAA GTC GCG CTT ACT GTA GAA GGC TAT CAA AGT AGC GGA AGT GCT	576		
Tyr Glu Val Ala Leu Thr Val Glu Gly Tyr Gln Ser Ser Gly Ser Ala			
180	185	190	
AAT GTA TAT AGC AAT ACA CTA AGA ATT AAC GGT AAC CCT CTC TCA ACT	624		
Asn Val Tyr Ser Asn Thr Leu Arg Ile Asn Gly Asn Pro Leu Ser Thr			
195	200	205	
ATT AGT AAT GAC GAG AGC ATA ACT TTG GAT AAA AAC AAT	663		
Ile Ser Asn Asp Glu Ser Ile Thr Leu Asp Lys Asn Asn			
210	215	220	

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH:** 221 amino acids
- (B) TYPE:** amino acid
- (D) TOPOLOGY:** linear

5

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal fragment

(vi) ORIGINAL SOURCE:

- (A) ORGANISM:** Bacillus
- (B) STRAIN:** 720/1

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Gln Ile Val Thr Asp Asn Ser Ile Gly Asn His Asp Gly Tyr Asp Tyr
1 5 10 15

Glu Phe Trp Lys Asp Ser Gly Gly Ser Gly Thr Met Ile Leu Asn His
20 25 30

Gly Gly Thr Phe Ser Ala Gln Trp Asn Asn Val Asn Asn Ile Leu Phe
35 40 45

Arg Lys Gly Lys Phe Asn Glu Thr Gln Thr His Gln Gln Val Gly
50 55 60

Asn Met Ser Ile Asn Tyr Gly Ala Asn Phe Gln Pro Asn Gly Asn Ala
65 70 75 80

Tyr Leu Cys Val Tyr Gly Trp Thr Val Asp Pro Leu Val Glu Tyr Tyr
85 90 95

Ile Val Asp Ser Trp Gly Asn Trp Arg Pro Pro Gly Ala Thr Pro Lys
100 105 110

Gly Thr Ile Thr Val Asp Gly Gly Thr Tyr Asp Ile Tyr Glu Thr Leu
115 120 125

Arg Val Asn Gln Pro Ser Ile Lys Gly Ile Ala Thr Phe Lys Gln Tyr
130 135 140

Trp Ser Val Arg Arg Ser Lys Arg Thr Ser Gly Thr Ile Ser Val Ser
145 150 155 160

Asn His Phe Arg Ala Trp Glu Asn Leu Gly Met Asn Met Gly Lys Met
165 170 175

Tyr Glu Val Ala Leu Thr Val Glu Gly Tyr Gln Ser Ser Gly Ser Ala
180 185 190

Asn Val Tyr Ser Asn Thr Leu Arg Ile Asn Gly Asn Pro Leu Ser Thr
195 200 205

Ile Ser Asn Asp Glu Ser Ile Thr Leu Asp Lys Asn Asn
210 215 220

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 744 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGAGACAAA AGAAATTGAC GTTGATTTA GCCTTTTAG TTTGTTTGC ACTAACCTTA 60
CCTGCAGAAA TAATTCAAGGC ACAAAATCGTC ACCGACAATT CCATTGGCAA CCACGATGGC 120
TATGATTATG AATTTGGAA AGATAGCGGT GGCTCTGGGA CAATGATTCT CAATCATGGC 180
GGTACGTTCA GTGCCAATG GAACAATGTT AACAAACATAT TATTCCGTAA AGGTAAAAAA 240
TTCAATGAAA CACAAACACA CCAACAAGTT GGTAACATGT CCATAAACTA CGGAGCCAAC 300
TTCCAACCAA ATGGTAATGC GTATTTATGC GTCTATGGTT GGACTGTTGA CCCTCTTGC 360
GAATATTATA TTGTCGACAG TTGGGGCAAC TGGCGTCCAC CAGGAGCAAC GCCTAAGGGG 420
ACCATCACTG TTGATGGAGG AACATATGAT ATCTACGAGA CTCTTAGAGT CAATCAACCC 480
TCCATTAAGG GGATTGCCAC ATTTAAACAA TATTGGAGTG TTCGAAGATC GAAACGCACG 540
AGTGGCACGA TTTCTGTCAG CAACCACTTT AGAGCGTGGG AAAACTTAGG GATGAATATG 600

GGGAAAATGT ATGAAGTCGC GCTTACTGTA GAAGGCTATC AAAGTAGCGG AAGTGCTAAT	660
GTATATAGCA ATACACTAAG AATTAACGGT AACCCCTCTCT CAACTATTAG TAATGACGAG	720
AGCATAACTT TGGATAAAAA CAAT	744

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 744 base pairs
- (B) TYPE: nucleic acid
- 5 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATG AGA CAA AAG AAA TTG ACG TTG ATT TTA GCC TTT TTA GTT TGT TTT	48	
Met Arg Gln Lys Lys Leu Thr Leu Ile Leu Ala Phe Leu Val Cys Phe		
-25	-20	-15

GCA CTA ACC TTA CCT GCA GAA ATA ATT CAG GCA CAA ATC GTC ACC GAC	96		
Ala Leu Thr Leu Pro Ala Glu Ile Ile Gln Ala Gln Ile Val Thr Asp			
-10	-5	-1 1	5

AAT TCC ATT GGC AAC CAC GAT GGC TAT GAT TAT GAA TTT TGG AAA GAT	144	
Asn Ser Ile Gly Asn His Asp Gly Tyr Asp Tyr Glu Phe Trp Lys Asp		
10	15	20

AGC GGT GGC TCT GGG ACA ATG ATT CTC AAT CAT GGC GGT ACG TTC AGT1	92	
Ser Gly Gly Ser Gly Thr Met Ile Leu Asn His Gly Gly Thr Phe Ser		
25	30	35

GCC CAA TGG AAC AAT GTT AAC AAC ATA TTA TTC CGT AAA GGT AAA AAA Ala Gln Trp Asn Asn Val Asn Asn Ile Leu Phe Arg Lys Gly Lys Lys	240
40 45 50	
TTC AAT GAA ACA CAA ACA CAC CAA GTT GGT AAC ATG TCC ATA AAC Phe Asn Glu Thr Gln Thr His Gln Gln Val Gly Asn Met Ser Ile Asn	288
55 60 65	
TAC GGA GCC AAC TTC CAA CCA AAT GGT AAT GCG TAT TTA TGC GTC TAT Tyr Gly Ala Asn Phe Gln Pro Asn Gly Asn Ala Tyr Leu Cys Val Tyr	336
70 75 80 85	
GGT TGG ACT GTT GAC CCT CTT GTC GAA TAT TAT ATT GTC GAC AGT TGG Gly Trp Thr Val Asp Pro Leu Val Glu Tyr Tyr Ile Val Asp Ser Trp	384
90 95 100	
GGC AAC TGG CGT CCA CCA GGA GCA ACG CCT AAG GGG ACC ATC ACT GTT Gly Asn Trp Arg Pro Pro Gly Ala Thr Pro Lys Gly Thr Ile Thr Val	432
105 110 115	
GAT GGA GGA ACA TAT GAT ATC TAC GAG ACT CTT AGA GTC AAT CAA CCC Asp Gly Gly Thr Tyr Asp Ile Tyr Glu Thr Leu Arg Val Asn Gln Pro	480
120 125 130	
TCC ATT AAG GGG ATT GCC ACA TTT AAA CAA TAT TGG AGT GTT CGA AGA Ser Ile Lys Gly Ile Ala Thr Phe Lys Gln Tyr Trp Ser Val Arg Arg	528
135 140 145	
TCG AAA CGC ACG AGT GGC ACG ATT TCT GTC AGC AAC CAC TTT AGA GCG Ser Lys Arg Thr Ser Gly Thr Ile Ser Val Ser Asn His Phe Arg Ala	576
150 155 160 165	
TGG GAA AAC TTA GGG ATG AAT ATG GGG AAA ATG TAT GAA GTC GCG CTT Trp Glu Asn Leu Gly Met Asn Met Gly Lys Met Tyr Glu Val Ala Leu	624
170 175 180	

ACT GTA GAA GGC TAT CAA AGT AGC GGA AGT GCT AAT GTA TAT AGC AAT 672
Thr Val Glu Gly Tyr Gln Ser Ser Gly Ser Ala Asn Val Tyr Ser Asn
185 190 195

ACA CTA AGA ATT AAC GGT AAC CCT CTC TCA ACT ATT AGT AAT GAC GAG 720
Thr Leu Arg Ile Asn Gly Asn Pro Leu Ser Thr Ile Ser Asn Asp Glu
200 205 210

AGC ATA ACT TTG GAT AAA AAC AAT 744
Ser Ile Thr Leu Asp Lys Asn Asn
215 220

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
(B) TYPE: amino acid
5 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal fragment

(vi) ORIGINAL SOURCE:

- 10 (A) ORGANISM: *Bacillus*
(B) SOURCE: 720/1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Arg Gln Lys Lys Leu Thr Leu Ile Leu Ala Phe Leu Val Cys Phe
-25 -20 -15

Ala Leu Thr Leu Pro Ala Glu Ile Ile Gln Ala Gln Ile Val Thr Asp
-10 -5 -1 1 5

Asn Ser Ile Gly Asn His Asp Gly Tyr Asp Tyr Glu Phe Trp Lys Asp
10 15 20

Ser Gly Gly Ser Gly Thr Met Ile Leu Asn His Gly Gly Thr Phe Ser
25 30 35

Ala Gln Trp Asn Asn Val Asn Asn Ile Leu Phe Arg Lys Gly Lys Lys
40 45 50

Phe Asn Glu Thr Gln Thr His Gln Gln Val Gly Asn Met Ser Ile Asn
55 60 65

Tyr Gly Ala Asn Phe Gln Pro Asn Gly Asn Ala Tyr Leu Cys Val Tyr
70 75 80 85

Gly Trp Thr Val Asp Pro Leu Val Glu Tyr Tyr Ile Val Asp Ser Trp
90 95 100

Gly Asn Trp Arg Pro Pro Gly Ala Thr Pro Lys Gly Thr Ile Thr Val
105 110 115

Asp Gly Gly Thr Tyr Asp Ile Tyr Glu Thr Leu Arg Val Asn Gln Pro
120 125 130

Ser Ile Lys Gly Ile Ala Thr Phe Lys Gln Tyr Trp Ser Val Arg Arg
135 140 145

Ser Lys Arg Thr Ser Gly Thr Ile Ser Val Ser Asn His Phe Arg Ala
150 155 160 165

Trp Glu Asn Leu Gly Met Asn Met Gly Lys Met Tyr Glu Val Ala Leu
170 175 180

Thr Val Glu Gly Tyr Gln Ser Ser Gly Ser Ala Asn Val Tyr Ser Asn
185 190 195

Thr Leu Arg Ile Asn Gly Asn Pro Leu Ser Thr Ile Ser Asn Asp Glu
200 205 210

Ser Ile Thr Leu Asp Lys Asn Asn
215 220

(2) INFORMATION FOR SEQ ID NO:7:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 81 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGAGACAAA AGAAATTGAC GTTGATTTA GCCTTTTAG TTTGTTTGC ACTAACCTTA
60
CCTGCAGAAA TAATTCAGGC A
81

(2) INFORMATION FOR SEQ ID NO:8:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 81 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATG AGA CAA AAG AAA TTG ACG TTG ATT TTA GCC TTT TTA GTT TGT TTT
Met Arg Gln Lys Lys Leu Thr Leu Ile Leu Ala Phe Leu Val Cys Phe
1 5 10 15

48

GCA CTA ACC TTA CCT GCA GAA ATA ATT CAG GCA
Ala Leu Thr Leu Pro Ala Glu Ile Ile Gln Ala
20 25

81

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
(B) TYPE: amino acid
5 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal fragment

(vi) ORIGINAL SOURCE:

- 10 (A) ORGANISM: Bacillus
(B) SOURCE: 720/1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Arg Gln Lys Lys Leu Thr Leu Ile Leu Ala Phe Leu Val Cys Phe
1 5 10 15

Ala Leu Thr Leu Pro Ala Glu Ile Ile Gln Ala
20 25

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 1513 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AAATTGAATT GTGTATATCT AATGATAACG ACAAATCGTC ACTGTTTTA AACTAATCTC	60
AAACCAATAC TTCTTTATTT AACGCTAACCC ACTTGCAATC TTATCACAAG AACATTCTTT	120
ATAGGAACCTT TCCCATTGTC AAGACGATAAA AAAATCTTTT TCCCCTATTT TATCTTATCG	180
CCTTGATCGG TTTAATTGT AAACTTTATT TTAGTTACG TGATGTTCCC TCATTACATAC	240
CATTAATCAC AGTTAACGCT AGAGTCATCT TTTTCGGTT CTCAAAAATA CCTGAAGAAC	300
ATTTATGTCA TATTTCTCA CGCCGCTCCA TAATGGAATA TATATACTCT TTTATACATA	360
TTAAGTAAAT TAGTATATAC TTGCGTTATC AAAATGTGAG ATAATCTAAT TGATCAAACA	420
AGCAGCTATC CAAAAAACAC TGATGTTGAC CTCTTAAAGA AGTGTCACTA TCTATGAAAA	480
GATAATTATC CAGTTCAAA ATTTGAAATA GTGTGTATGG AATAGTTTGA ATGTCAACTG	540
CTGTGAAAGG AGGGTAGGTA GTACCGTAGA CTTCATTACC AAAAATTAGT TGTAAAAAAA	600
TTAAAAGGAG GAATGCCTAA TGAGACAAAA GAAATTGACCG TTGATTTAG CCTTTTIACT	660
TTGTTTGCA CTAACCTTAC CTGCAGAAAT AATTCAAGGCA CAAATCGTCA CCGACAATT	720
CATTGGCAAC CACGATGGCT ATGATTATGA ATTTGGAAA GATAGCGGTG GCTCTGGAC	780
AATGATTCTC AATCATGGCG GTACGTTCAAG TGCCCAATGG ACAATGTTA ACAACATATT	840
ATTCCGTAAA GGTAAAAAT TCAATGAAAC ACAAACACAC CAACAAGTTG GTAACATGTC	900
CATAAAACTAC GGAGCCAATC TCCAACCAAA TGGTAATGCG TATTTATGCG TCTATGGTTG	960
GACTGTTGAC CCTCTTGTG AATATTATAT TGTGACAGT TGGGGCAACT GGCGTCCACC	1020
AGGAGCAACG CCTAAGGGGA CCATCACTGT TGATGGAGGA ACATATGATA TCTACGAGAC	1080
TCTTAGAGTC AATCAACCCT CCATTAAGGG GATTGCCACA TTTAAACAAT ATTGGAGTGT	1140

TCGAAGATCG AAACGCACGA GTGGCACGAT TTCTGTCAGC AACCACTTA GAGCGTGGGA	1200
AAACTTAGGG ATGAATATGG GGAAAATGTA TGAAGTCGCG CTTACTGTAG AAGGCTATCA	1260
AAGTAGCGGA AGTGCTAATG TATATAGCAA TACACTAAGA ATTAACGGTA ACCCTCTCTC	1320
AACTATTAGT AATGACGAGA GCATAACTTT GGATAAAAAC ATTAAAAAT CCTTATCTCT	1380
TTCGGTTCAG TTCTCATTAT TTTCAAATAA CCTCCGGTT GGATCTTTTC CAACGGGAGG	1440
TTTTATTGGA AAGGTTAAGT ATAGTATACT CCGATTCCAT CCAGAGGAAT GCTTGAAACA	1500
CCTCCGTCAC TAG	1513

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1513 base pairs
- (B) TYPE: nucleic acid
- 5 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

AAATTGAATT GTGTATATCT AATGATAACG ACAAAATCGTC ACTGTTTTA AACTAATCTC 60

AAACCAATAAC TTCTTTATTT AACGCTAACCC ACTTGCAATC TTATCACAAG AACATTCTTT 120
ATAGGAACCTT TCCCATTGTC AAGACGATAA AAAATCTTTT TCCCCTATTT TATCTTATCG 180
CCTTGATCGG TTTAATTGT AAACTTTATT TTAGTTACG TGATGTTCCC TCATTCATAC 240
CATTAATCAC AGTTAACGCT AGAGTCATCT TTTTCGGTT CTCAAAAATA CCTGAAGAAC 300
ATTTATGTCA TATTTCTCA CGCCGCTCCA TAATGGAATA TATATACTCT TTTATACATA 360
TTAAGTAAAT TAGTATATAC TTGCGTTATC AAAATGTGAG ATAATCTAAT TGATCAAACA 420
AGCAGCTATC CAAAAAACAC TGATGTTGAC CTCTTAAAGA AGTGTCACTA TCTATGAAAA 480
GATAATTATC CAGTTCAAA ATTTGAAATA GTGTGTATGG AATAGTTGA ATGTCAACTG 540
CTGTGAAAGG AGGGTAGGTA GTACCGTAGA CTTCATTACC AAAAATTAGT TGTAAAAAAA 600
TTAAAAGGAG GAATGCCTA ATG AGA CAA AAG AAA TTG ACG TTG ATT TTA GCC 652
Met Arg Gln Lys Lys Leu Thr Leu Ile Leu Ala
-25 -20
TTT TTA GTT TGT TTT GCA CTA ACC TTA CCT GCA GAA ATA ATT CAG GCA 700
Phe Leu Val Cys Phe Ala Leu Thr Leu Pro Ala Glu Ile Ile Gln Ala
-15 -10 -5 -1
CAA ATC GTC ACC GAC AAT TCC ATT GGC AAC CAC GAT GGC TAT GAT TAT 748
Gln Ile Val Thr Asp Asn Ser Ile Gly Asn His Asp Gly Tyr Asp Tyr
1 5 10 15
GAA TTT TGG AAA GAT AGC GGT GGC TCT GGG ACA ATG ATT CTC AAT CAT 796
Glu Phe Trp Lys Asp Ser Gly Gly Ser Gly Thr Met Ile Leu Asn His
20 25 30

GGC GGT ACG TTC AGT GCC CAA TGG AAC AAT GTT AAC AAC ATA TTA TTC Gly Gly Thr Phe Ser Ala Gln Trp Asn Asn Val Asn Asn Ile Leu Phe	35	40	45	844
CGT AAA GGT AAA AAA TTC AAT GAA ACA CAA ACA CAC CAA CAA GTT GGT Arg Lys Gly Lys Lys Phe Asn Glu Thr Gln Thr His Gln Gln Val Gly	50	55	60	892
AAC ATG TCC ATA AAC TAC GGA GCC AAC TTC CAA CCA AAT GGT AAT GCG Asn Met Ser Ile Asn Tyr Gly Ala Asn Phe Gln Pro Asn Gly Asn Ala	65	70	75	940
TAT TTA TGC GTC TAT GGT TGG ACT GTT GAC CCT CTT GTC GAA TAT TAT Tyr Leu Cys Val Tyr Gly Trp Thr Val Asp Pro Leu Val Glu Tyr Tyr	85	90	95	988
ATT GTC GAC AGT TGG GGC AAC TGG CGT CCA CCA GGA GCA ACG CCT AAG Ile Val Asp Ser Trp Gly Asn Trp Arg Pro Pro Gly Ala Thr Pro Lys	100	105	110	1036
GGG ACC ATC ACT GTT GAT GGA GGA ACA TAT GAT ATC TAC GAG ACT CTT Gly Thr Ile Thr Val Asp Gly Gly Thr Tyr Asp Ile Tyr Glu Thr Leu	115	120	125	1084
AGA GTC AAT CAA CCC TCC ATT AAG GGG ATT GCC ACA TTT AAA CAA TAT Arg Val Asn Gln Pro Ser Ile Lys Gly Ile Ala Thr Phe Lys Gln Tyr	130	135	140	1132
TGG AGT GTT CGA AGA TCG AAA CGC ACG AGT GGC ACG ATT TCT GTC AGC Trp Ser Val Arg Arg Ser Lys Arg Thr Ser Gly Thr Ile Ser Val Ser	145	150	155	1180
AAC CAC TTT AGA GCG TGG GAA AAC TTA GGG ATG AAT ATG GGG AAA ATG Asn His Phe Arg Ala Trp Glu Asn Leu Gly Met Asn Met Gly Lys Met	165	170	175	1228

TAT GAA GTC GCG CTT ACT GTA GAA GGC TAT CAA AGT AGC GGA AGT GCT 1276
Tyr Glu Val Ala Leu Thr Val Glu Gly Tyr Gln Ser Ser Gly Ser Ala
180 185 190

AAT GTA TAT AGC AAT ACA CTA AGA ATT AAC GGT AAC CCT CTC TCA ACT 1324
Asn Val Tyr Ser Asn Thr Leu Arg Ile Asn Gly Asn Pro Leu Ser Thr
195 200 205

ATT AGT AAT GAC GAG AGC ATA ACT TTG GAT AAA AAC AAT TAAAAATCCT 1373
Ile Ser Asn Asp Glu Ser Ile Thr Leu Asp Lys Asn Asn
210 215 220

TATCTCTTTC GGTCAGTTC TCATTATTT CAAATAACCT CCCGGTTGGA TCTTTTCCAA 1433

CGGGAGGTTT TATTGGAAAG GTTAAGTATA GTATACTCCG ATTCCATCCA GAGGAATGCT 1493

TGAAACACCT CCGTCACTAG 1513

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 619 base pairs
- (B) TYPE: nucleic acid
- 5 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AAATTGAATT GTGTATATCT AATGATAACG ACAAAATCGTC ACTGTTTTTA AACTAATCTC 60
AAACCAATAC TTCTTTATTT AACGCTAACCC ACTTGCAATC TTATCACAAG AACATTCTTT

ATAGGAACCTT TCCCATTGTC AAGACGATAA AAAATCTTTT TCCCCTATTT TATCTTATCG 180

CCTTGATCGG TTTAATTGT AAACTTTATT TTAGTTACG TGATGTTCCC TCATTCATAC	240
CATTAATCAC AGTTAACGCT AGAGTCATCT TTTTCGGTT CTCAAAAATA CCTGAAGAAC	300
ATTTATGTCA TATTTCTCA CGCCGCTCCA TAATGGAATA TATATACTCT TTTATACATA	360
TTAAGTAAAT TAGTATATAC TTGCGTTATC AAAATGTGAG ATAATCTAAT TGATCAAACA	420
AGCAGCTATC CAAAAAACAC TGATGTTGAC CTCTTAAAGA AGTGTCACTA TCTATGAAAA	480
GATAATTATC CAGTTCAAA ATTTGAAATA GTGTGTATGG AATAGTTGAA ATGTCAACTG	540
CTGTGAAAGG AGGGTAGGTA GTACCGTAGA CTTCATTACC AAAAATTAGT TGTAAAAAAA	600
TTAAAAGGAG GAATGCCTA	619

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 150 base pairs
- (B) TYPE: nucleic acid
- 5 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TAAAAATCCT TATCTTTTC GGTTCAAGTTC TCATTATTTT CAAATAACCT CCCGGTTGGA	60
TCTTTTCCAA CGGGAGGTTT TATTGGAAAG GTTAAGTATA GTATACTCCG ATTCCATCCA	120
GAGGAATGCT TGAAACACCT CCGTCACTAG	150

(2) INFORMATION FOR SEQ ID NO:14:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCCCCCTACG TAGCGGCCGC CCCGGCCGGT AACCTAGGAA GTCAGCGCCC TGCACC

56

(2) INFORMATION FOR SEQ ID NO:15:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCCCCCTACG TAGGCCGGGG CGGCCGCGGT TACCTAGGGC CTCGTGATAC GCCTAT

56

(2) INFORMATION FOR SEQ ID NO:16:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ACGAGGAAAG ATGCTGTTCT TGTAAATGAG T

31

(2) INFORMATION FOR SEQ ID NO:17:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TACCTTGTCT ACAAACCCC

19

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- 15
- (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGGTCGCCGC ATACACTA

18

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- 5 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCCCCCCCCCG GTAACCTGCA TTAATGAATC GGCCAA

36

10 (2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- 15 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCCCCCCCCCG GTTACCGTAT TTATTAACCTT CTCCTAGTA

39

(2) INFORMATION FOR SEQ ID NO:21:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCCCCCTCTA GATTAATTAA CCAAGCTTGG GATCCGTCGA CCTGCAGATC

50

(2) INFORMATION FOR SEQ ID NO:22:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCCCCCTGAA ATCAGCTGGA CTAAAAGGGA TGCAATTTC

39

(2) INFORMATION FOR SEQ ID NO:23:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCCCCGTCG ACCGCATGCG CGGGCACAGC

30

(2) INFORMATION FOR SEQ ID NO:24:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCCCCCGCAT GCGCAAATCG TCACCGACAA TTCCATTGG

39

15 (2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: nucleic acid (synthetic oligo-nucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TACCTTGTCT ACAAACCCC

19

(2) INFORMATION FOR SEQ ID NO:26:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 185 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TCATGTAACT CGCCTTGATC TATTTCAATT GTATCAAAGG ATTTATACAC AAACAAGAGA	60
CATCCATGCC GGGTTAAAGC AGTATCGTTC CATCTAACAG AGAAGGNCTG CATGAAAGGA	120
GGTGATGGGT TTTTCATCTT AGGGATGACA GAACAATACG GATGAAAAAA GGAGAGGGAT	180
GGAAA	

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 81 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATGAATTGAA AAAGATTGAG GCTGTTGTT GTGATGTGTA TTGGATTGT GCTGACACTG	60
ACGGCTGTGC CGGCTCATGC G	81

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 81 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATG AAT TTG AAA AGA TTG AGG CTG TTG TTT GTG ATG TGT ATT GGA TTT	48
Met Asn Leu Lys Arg Leu Arg Leu Leu Phe Val Met Cys Ile Gly Phe	
1 5 10 15	

GTG CTG ACA CTG ACG GCT GTG CCG GCT CAT GCG
Val Leu Thr Leu Thr Ala Val Pro Ala His Ala
20 25

81

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- 5 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Asn Leu Lys Arg Leu Arg Leu Leu Phe Val Met Cys Ile Gly Phe
1 5 10 15
Val Leu Thr Leu Thr Ala Val Pro Ala His Ala
20 25